**Lab Virtual Screen 3: Setting up the protein file in GOLD/HERMES**

NOTE: this lab is not due for 2 weeks

**OBJECTIVE:** The purpose of this lab is to screen a library of compounds for their ability to bind the active site of a phosphatase enzyme as predicted by the GOLD docking program. In the past, you have been supplied the protein file. In this lab, you will prepare the file yourself and then dock 5 Control compounds along with a library of novel ligands.

**INTRODUCTION**

**Disease and Target:** Bacterial and protozoan parasites need to communicate signals in order to infect and survive in a host cell. These messages can be passed within the pathogen cell itself or may take place between the pathogen and the host cell. Many diseases such as tularemia, tuberculosis and the bubonic plague rely on this strategy to render host macrophages susceptible to infection. While not a classical target for inhibition therapy, these pathways may be a new tactic for anti-infectives that could be coupled with current treatments to circumvent the rise of anti-bacterial resistance.

The PTP enzymes are responsible for dephosphorylating proteins involved in some of these signaling pathways. Phosphorylation of client proteins is a common post translational modification in eukaryotic cells which acts as a reversible ‘switch’ to permit signal transduction or biological activity. Kinases add phosphate groups while phosphatases remove them. Phosphorylation can take place on amino acids which contains an hydroxyl (-OH) group on the side chain: serine, threonine, tyrosine and sometimes on aspartate and even one of the nitrogens in the imidazole ring of histidine. PTPs recognize tyrosine phosphate groups for subsequent removal as shown below.

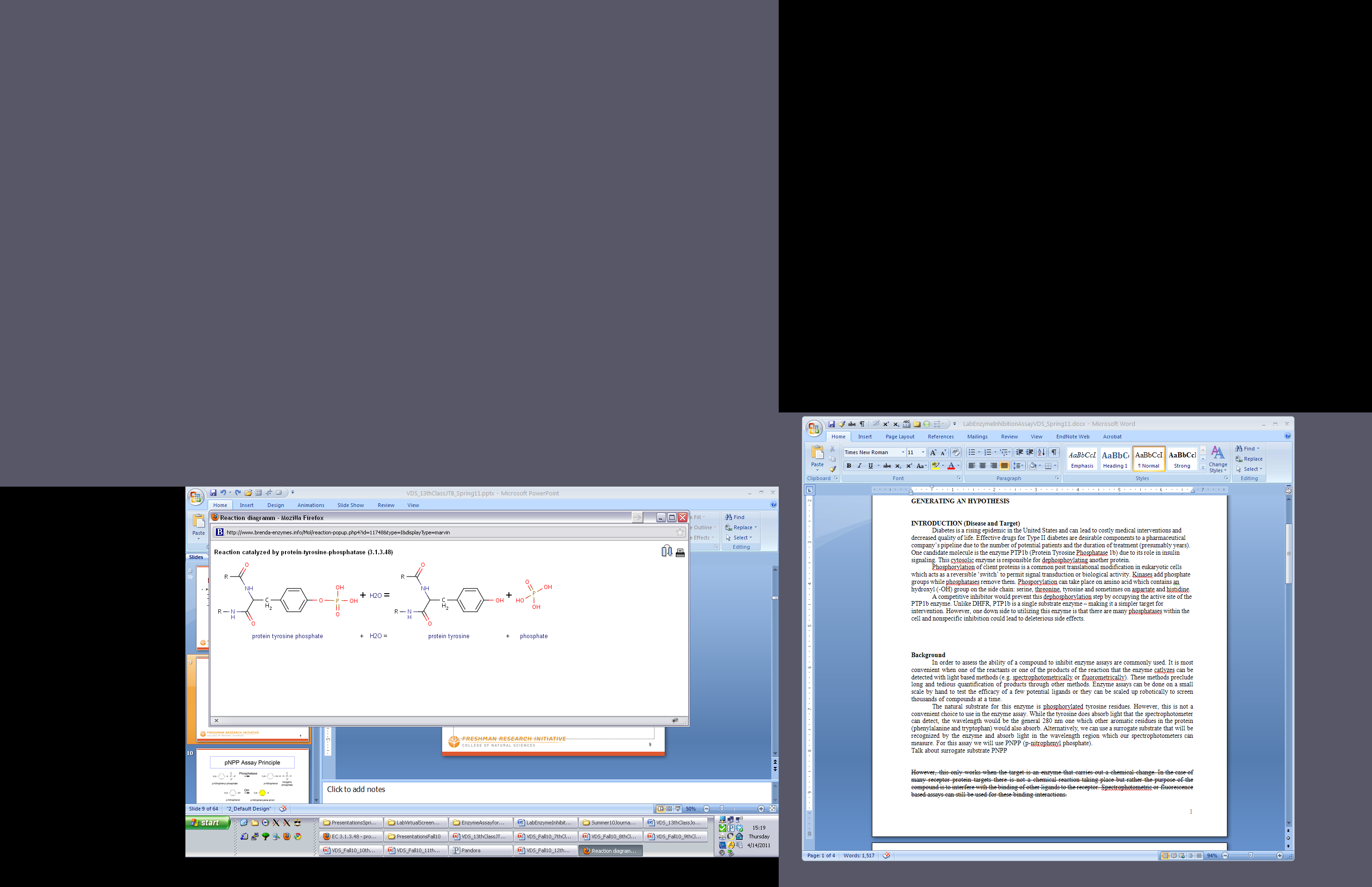


Figure 1. Reaction catalyzed by protein-tyrosine-phosphatase (3.1.3.48) upon natural substrate

A competitive inhibitor to PTPs would prevent this dephosphorylation step by occupying the active site of the enzyme. Unlike DHFR, PTPs are single substrate enzymes – making them a simpler target for intervention. One down side to utilizing this enzyme as a therapeutic pathway is that there are many phosphatases within the cell. Any drug candidate would have to be carefully monitored for nonspecific inhibition which could lead to deleterious side effects.

We have studied phosphatases extensively for many of our targets in VDS. Phosphatases are utilized in cell-cell communication as well as intracellular communication. Therefore they can be important for the viability and/or infectivity of an organism. For example, some pathogens secrete phosphatases into macrophages to disrupt cell signaling pathways and prevent them from killing the invading pathogen cell. The infectious agent can then take up residence inside the macrophage as the host. Phosphatases also happen to be easy to assay in the wet lab since we can often use spectrophotometric surrogate substrates. These proteins are usually single substrate enzymes which make for more facile screening assays. The library will contain predominantly novel compounds that have not been tested against these structures in the wet lab. There will also be 2 known inhibitors of phosphatases taken from the BindingDB.com database as well as one compound that is used as a surrogate substrate in our enzyme assays to detect the activity. Also, aspirin (acetylsalicylic acid) will be docked to see if this common drug has any binding implications for the phosphatase.

**VIRTUAL SCREENING**

* 1st Run
  + Screen ~1,000 ligands at 0.1 autoscale on **5** processors
  + Save 10% = ~ 100
* 2nd Run
  + Screen the ~100 ligands at 2 autoscale on **3** processors
  + Save 10 per processor = 30 final ligands
* Analyze Bestranking list
* Examine poses in PyMol

In your lab notebook, record your steps and which library and ligands that you have screened. It is particularly important to keep track of your filenames, ligands and proteins and how you organized them.

**DOWNLOAD THE CRYSTAL STRUCTURE**:

For this lab, you will get to pick which crystal structure you would like to use. So, let’s go find a phosphatase of an organism for which we are interested in developing a therapeutic drug.

In the mini-research write ups – you chose infectious diseases that were of interest. You could use this same disease organism or pick a new one. Pick a bacterial or protozoan disease to find an enzyme to target (viral organisms use these phosphatases less frequently)

Go to **PDB website**, do the Advanced search (on right hand side)

Choose Enzyme Classification Browser (Open Pop-up)

Click on the Down Arrow to expand and Search under **Hydrolases**

---Note: if you accidentally click on the Name instead of the Down Arrow, close the pop-up and re-open it

Click on the Down Arrow on the one for ‘**Acting on ester bonds**’

Click on the Down Arrow on the one for ‘**Phosphoric Monoester Hydrolases’**

Click on the ‘**Protein-Tyrosine-Phosphatase’**

* **Paste in your lab notebook a picture of an ester bond**

**Now,** let’s add another Restriction to the search. It is usually considered best to only do virtual screening using crystal structures that have a resolution of less than 3 Angstroms. The higher the resolution, the more confident you are that any given atom is actually where the PDB says it is in space.

Choose to **Add Search Criteria** then find the **X-ray Resolution** in the pull down list.

Enter **‘0’** for Min and ‘**3’** for the Max resolution.

Click on **Result Count**

Then **Submit Query** to search for both of the search qualifications.

Now, you should see a list with at least a few hundred entries. To further narrow our research study, lets only choose one classification under **Taxonomy**.

If you want to look at a bacterial target – choose the ‘**Bacteria’** under the **Taxonomy** tab. And then choose one that sounds interesting to you and satisfied the criteria below and that hasn’t been taken by someone else

Select one that satisfies the following criteria:

* + Has only 1 inhibitor present in the active site (cannot be allosteric either)
    - NOTE: It is ok if it has some other ‘ligands’ in solution around the protein – just as long as they are not in the active site. Other molecules that may be in solution but are not really ‘ligand’ are: Sodium, Potassium, Glycerol, Tris. (PO4 counts as a ligand – it is the phosphate that is cleaved in the reaction).
    - TIP: Use PyMol or the built-in Jmol viewer to verify where a ligand is in the structure.
  + Choose a file that has no mutations in the amino acid sequence of the active site
    - It is ok to have mutations peripherally though
    - Best if it does not contain **Modified Residues** (i.e. MSE (selenomethionine))

Alternatively: If you want to look at a protozoan parasite – select **Eukaryota** here since protozoans have nuclei. Under **Organism** – find a protozoan (you will need to Wiki a few if you don’t know). Check the ‘**Other’** selection to see if those are protozoan too.

**DOWNLOAD THE PROTEIN –**

**write down what organism you chose and the disease to which it is implicated.**

On the rhs (right hand side) of the PDB page, there is a link to the text file from the ‘right click’ menu.

Download the file to your Desktop (or wherever you want)

**DOWNLOAD THE LIGAND**

Also, download the PubChem version of the ligand (don’t download the PDB version – it isn’t properly configured). In the PDB page, go to **Ligand Chemical Component**. Click on the Identifier for the ligand (e.g. PSY) and this will take you to the **Ligand Summary** page. Now, follow the PubChem link to download the structure as 3D by clicking on the ‘**SDF´**icon in the top right of the PubChem page. Choose to save as a **3D SDF.** Keep the name given.

NOTE: this ligand may not be fully minimized. We will actually skip the ligand minimization step for the sake of time/simplicity.

THE PubChem site states this about their 3-D ligands:

“Each provided theoretical 3-D conformer is not at an energy minimum and may not represent the lowest energetic form in vacuum, solvent, or a binding pocket. Rather, the theoretical 3-D description consists of low energy conformers selected from a conformer model (a description of the conformational flexibility of a chemical structure consisting of multiple 3-D representations or poses sampled using average atom pair- wise RMSD {root mean squared distance} threshold) describing energetically-accessible and (potentially) biologically relevant conformations of a chemical structure.”

NOTE: if there is not a PubChem version of the ligand, then it is ok to use the PDB version.

**VERIFY AND FIX STRUCTURES IN PYMOL**:

Open the ligand in PyMol and make sure it looks legit **Q:** is it in 3D? Does it have Hydrogens?

Now, open the protein file in PyMol. IF there is more than one chain present we will want to remove them. Open your structure in PyMol, and Color ‘**By Chain’** You should see each of the chains as a different color. Now, change your ‘**Selecting’** mode to **‘Chains’** click on the chain you want to get rid of (usually B, C, or D – but make sure the one you keep has the ligand present in the active site) . When it is made as a new selection by PyMol – **Remove Atoms**

**Then go to File >> Save Molecule >> as**

**<the pdb identifier>chainA.pdb for example: XXXXchainA.pdb or XXXXchainB.pdb**

**MOLPROBITY ANALYSIS OF PROTEIN**

Go to MolProbity website – TIP: use Internet Explorer

‘**Choose File**’ - Load **your PDB** file to the website (use the file you have from above instead of entering the PDB identifier directly into MolProbity)

Copy down the resolution (to put in your table below)

Add Hydrogens,

On next page, Choose the option to Add the flips ‘**Asn/Gln/His flips’**

Leave all the defaults checked and ‘**regenerate H’s’**

If it asks you to save the file with Hydrogens – skip it for now.

**Analyze all-atom contacts and geometry**

Use all defaults **>> Run Programs to Perform these Analyses**

**View the Multi-Criterion Chart**

Save the data for your table below

View the **Multi-Criterion Kinemage** to see what type of errors exist in the active site.

Make note of which type they are for your table below.

HINT: to see the active site – toggle on and off the **‘hets’** button

**Include** a ‘snip’ of your traffic light table from MolProbity (i.e. green, yellow, red table)

**Also include** these values for the structure in your notebook

|  |
| --- |
| **Resolution:** |
| **Error in Active Site:** |

**Q:** How good is the structure relative to the ones you used for the VS2 lab?

**Download** your PDB file with REMARK 40 WITHOUT Hydrogens – we will make GOLD add the H’s later

The filename should show up as something like:

XXXXchainAF.pdb (not XXXXchainAFH.pdb)

**TRANSFER PROTEIN AND LIGAND TO DDFE**

Login to the DDFE using WinSCP

ddfe.cm.utexas.edu

In virtual screening, it is important to keep your file structure organized and to reduce redundant files.

Create a directory structure on the DDFE like this below. You will be making 1 new directory for now.

e.g. /home/chem204/2013/YOURUTEID/**LabVS3**

Transfer the **protein** (.pdb) from your local computer over to your directory in the DDFE.

This should be the XXXchainAF.pdb

Transfer your 3D **ligand** file from PubChem over as well.

**LIGAND LIBRARY SELECTION**

In a drug discovery project, it is advantageous to find new and unique scaffolds that are good inhibitors. A scaffold is basically the ‘backbone’ of the ligand and constitutes its core physico-chemical characteristics, in particular its steric properties. Ideally, new scaffolds will have different pharmacological effects than existing drugs and may, therefore, offer alternative therapies. Consequently, the Chembridge diversity set of ligands was put together to satisfy a wide range of scaffolds. Ultimately, once some hits are found, the researcher can then optimize a scaffold by adding, changing or deleting moieties and atoms on the outside.

The Chembridge Diveristy set consists of approximately 50,000 ligands that are derived from many different scaffolds. This library has been divided up into blocks of around 5,000 for each of you to screen against our target: the phosphatase enzyme that you selected. In additions to these untested compounds, it is also important to include some known active compounds in a virtual screen so that we have a relative comparison of GOLD fitness scores.

**Select one of the 50 libraries to screen** from /home/chem204/DatabasesVDS/LabVS3\_PTP1bLibrary

**CB1k\_1.sdf CB1k\_2.sdf CB1k\_3.sdf CB1k\_4.sdf CB1k\_5.sdf**

**CB1k\_6.sdf CB1k\_7.sdf …etc. …………………………………………. CB1k\_50.sdf**

Use the **countsdf.pl** script in the LabVS3\_Library folder to find out how many ligands are present for the one you selected.

To run this script (sort of like a mini-program) type:

**$perl countsdf.pl**

Copy over your **CB1k\_X.sdf** library from the **/DatabasesVDS** to your **VS3** directory.

Also copy over the **countsdf.pl** to use later

**NOTE:** Normally we don’t copy over large libraries to our individual folders. But since we want to add some positive and negative control ligands to the library, we will need to do so.

**ADDING POSITIVE AND NEGATIVE CONTROLS TO LIBRARY**

We need to get some positive and negative conrols to add to the library. You will need to gather the physico-chemical properties (**Molecular weight, H-bond Donors, H-bond Acceptors, LogP**) for your final results table and also download the SDF files.

Go to the BindingDB database (<http://www.bindingdb.org/bind/index.jsp>) and under the **Full Search** type in ‘**phosphatase’**.

Scroll down to Target: **protein tyrosine phosphatase ptpb**

1. **POSITIVE CONTROL #1**

The first entry is a potential inhibitor for a phosphatase from *Mycobacterium tuberculosis*.

Make note of the **Ki** value in **nM**

Click on the **PC cid** and then download the structure as 3D by clicking on the ‘**SDF´**icon in the top right. Choose to save as **3D SDF** and record the **Physico-Chemical properties.**

* add the name ‘**pos1’** to the end of the file (e.g. CID\_ 44542081**pos1**.sdf)
* Open the file in WORDPAD (right click >> Open With) and add ‘**pos1’** to the end of the identifier in the 1st line

E.g. 44542081**pos1**

1. **POSITIVE CONTROL #2**

**Now go back and pick ANOTHER compound in the list for PTPb in Mycobacterium tuberculosis**

Make note of the **Ki** value in **nM**

Download this as a ‘3D’ SDF file and record the Physico-Chemical properties.

* add the name ‘**pos2’** to the end of the file & to the end of the 1st line as you did above

1. **POSITIVE CONTROL – surrogate substrate**

**pNPP** is a substrate that is used by researchers for phosphatases in the lab because it will absorb light in a spectrophotometer when it is hydrolyzed and can, therefore, be an indicator of enzyme activity.

GO to PubChem website and search for ‘**4-nitrophenyl phosphate’** (also called **pNPP**), down load the structure as 3D by clicking on the ‘**SDF´**icon in the top right. Chose to save as **3D SDF** and record the **Physico-Chemical properties.**

* add the name ‘**pNPP**’ to the end of the file & to the end of the 1st line as you did above

e.g. **CID\_378pNPP.sdf** and **378pNPP**

1. **RANDOM CONTROL**

Find **Aspririn** in the PubChem database and download it as **3D** **SDF** and record the **Physico-Chemical properties.**. We don’t have any evidence that aspirin will bind. So, it is a RANDOM or NEGATIVE CONTROL

* add the name ‘**aspirin’** to the end of the file & to the end of the 1st line as you did above

1. **Original Ligand that is in the PDB**

This is the ligand you downloaded on page 2 from PubChem.

Docking this is essentially a ‘validation dock’ to see if you docking program can properly dock a true ligand.

* add the name ‘**original’** to the end of the file & to the end of the 1st line as you did above

You should have 5 new compounds now from the PubChem database. Open each of them in PyMol to verify they look ‘ok’

**LIGAND PROTONATION STATE**

The ligands may or may not be in the correct protonation state when they are downloaded from the PubChem Database. Usually we will dock ligands at physiological pH – which is pH 7.4 (or 7.2).

There are 2 ways to do this

1. Go to OpenBabel on a Windows computer.

* Choose SDF as the INPUT FORMAT type, browse for your ligand file
* Enter **“7.4”** to ‘Add hydrogens appropriate for this pH’
* Output Format = SDF
* Name your output file in a useful manner. E.g. **CID\_378pNPP\_pH.sdf**
* **CONVERT**
* **Repeat this for each of your 5 Controls**
* Open each of them in PyMol to verify they look ‘ok’ – do the hydrogens change from the original?
* Transfer to the DDFE folder

1. Use command line version of Babel

* Transfer your files to the DDFE
* Open a Terminal Window in the folder where your files are
* Type this command to add Hydrogens appropriate for pH 7.4

$babel –isdf inputfilename.sdf –osdf outputfilename.sdf –p 7.4

e.g. $babel –isdf CID\_378pNPP.sdf –osdf CID\_378pNPP\_ph.sdf –p 7.4

* Open each of them in PyMol to verify they look ‘ok’– do the hydrogens change from the original?

NOTE: your ~1000 ligand library is already pH’d.

**CONCATENATING THE LIBRARY**

Now **append** all of these to the beginning of your Chembridge library.

$cat CID<POSCONTROL1\_pH>.sdf CID<POSCONTROL2\_pH>.sdf CID<378pNPP\_pH>..sdf CID<aspirin\_pH>.sdf CID<original\_pH>.sdf CB1k\_**#**.sdf >> **CB1kAll\_pH.sdf**

Where you fill in the real filenames in place of the brackets <>.

You are combining all of these files into one big one so that you can do virtual screening.

Your concatenated file should have the CB ligands + your 5 Control ligands

**CHECK YOUR LIGAND LIBRARY**

Use the **countsdf.pl** script to count the number of ligands. If you didn’t copy this script before – go get a copy in the /DatabasesVDS/LabVS3\_PTP1bLibrary directory and put it in your folder to use.

To run this script (sort of like a mini-program) type:

**$perl countsdf.pl**

**Write down the number of ligands in your library: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Transfer your newly made library over to your local computer. Open it in **PyMol** and check to be sure all the ligands with your Controls are present.

**IF,** your library failed to be assemble correctly, you may need to edit the Controls.

Open the Controls (Original, Pos1, Pos2, pNPP, Aspirin) in WORDPAD.

Check to see that there is only one empty line after the four dollar symbols ($$$$)

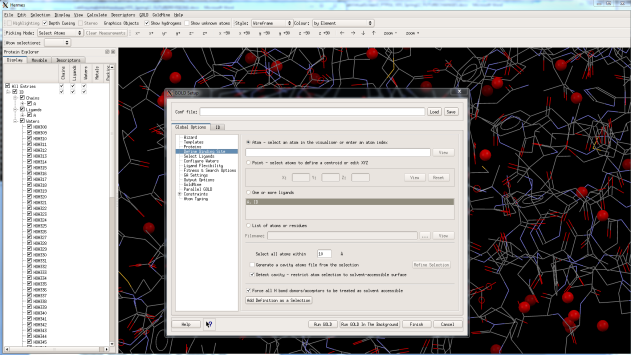
If there are two empty lines, remove the last empty line at the end of the file!

Otherwise, the files cannot be joined together properly!

**Q:** State an hypothesis about how you think these compounds will fare in the virtual screening of the phosphatase that you selected. Address the controls you are using and the novel ligands.

**PROTEIN PREPARATION**

When protein files are obtained from the Protein Data Bank, they are in more of a raw format. The PDB file needs to be cleaned up before a virtual screening run can be done on it. First, we checked the structure in MolProbity to determine the quality. Next in GOLD, we will add hydrogens to those atoms which need them so that hydrogen bond interactions can be more accurately assessed. Waters molecules will also be removed to speed up docking- unless we know of a certain water molecule that should stay. Lastly, the active site will be defined by using whichever ligand is currently in the structure from the X-ray crystallography experiment. Then this ligand will be extracted so that we can dock the new compounds into the active site in place of it.

We will use the visual interface (GUI – graphical user interface) of **GOLD** along with its companion program **Hermes** to carry out these preparation steps. We will walk through the protein set up check list and concurrently generate a **gold.conf** file which determines the parameters for our docking.

**NOTE:** these instructions are for Windoze. If you insist upon using your MacIntrash – see the Tips&Hints Page of our Wikispaces page for help.

**Connecting to the graphical interface for GOLD**

Make remote connection to DDFE using a graphical user interface (GUI) for GOLD

Open Xming server

Go to Start, Programs, Xming, Xming

Open Xlaunch

Go to Start, Programs, Xming, XLaunch

Figure 2. GOLD & Hermes

Select ‘Multiple Windows’

Select ‘Start no client’,

Skip next screen by selecting ‘Next’ then ‘Finish’ on next screen

Open **Putty** in Programs

Connect to Host Name: ddfe.cm.utexas.edu on Port 22 using SSH

On the left side of the window, Select the ‘SSH’ tab and then the ‘X11’ or ‘Tunnels’ tab

‘Enable X11 forwarding’

X display location: leave blank or enter localhost:0

# this is the default display on your computer

‘Open’

Login as user: type your user name for the DDFE (your UTEID)

Enter password

You must put yourself into the directory where your protein file is.

Type ‘ls’ to see the contents and ‘cd’ to change directories

Then type this to open gold with the graphical user interface:

**$gold**

Ignore the ‘BadFont’ error message, if present

Don’t load a Conf file at the top (that is what you will be making here)

Step through the Configuration Options to set up your file

Skip Wizard

Skip Templates

**Protein > Load protein** <nameofyourpdbfile>.pdb

Gold 5.0 has a separate window for Global Options and a specific window for operations on your protein.

**Under the tab for your PDB file name (to the right) – the tab may just say “ID”:**

**Protonation & Tautomers** > Add Hydrogens

Write down how many hydrogens added.

Skip - **Flip Asn GLn and HIS tautomers**. We won’t worry about these right now.

**Extract/Delete Waters**: ‘Delete Remaining Waters’ (don’t select any of them to save).

Write down how many waters removed.

**Delete Ligand**

NOTE: If there is more than one ligand, you will need to go into the **Hermes** visualizer window to figure out which ligand

Go to View >> Protein Explorer

Click on the ‘+’ (plus sign) to see the different objects.

Extract the Ligand, save as ‘**LigandExtracted.mol2’**

(this will be saved for defining the cavity site)

NOTE: you already saved a different PubChem version of this ligand for validation docking

**Back in WinSCP** – make sure your LigandExtracted has an extension

If not, then add it to the file (just add **.mol2** to the end)

Skip the remaining options for the protein.

**Under Global Options:**

Define Binding Site –‘Select One or more ligands’

‘One or more ligands’ - choose the single ligand that you had extracted.

‘Select all atoms within 7.5 Angstroms

Leave ‘Generate a cavity’ unchecked

Check – ‘Detect cavity’

Check – ‘Force all H bond donors/acceptors ….”

– verify active site in image on the Hermes visualizer

(only a small region around the ligand of the protein will be highlighted in gray)

In the Gold GUI – go back to **Global Options**

Select Ligands – you will need find where your library is (probably in your VS3 directory)

e.g. **CB1kAll\_pH.sdf**

This is the file you need to link to for your ligand library.

Then make sure the number of conformations per ligand or **GA Runs** is set to ‘10’

Skip the Reference Ligand

Skip ‘Configure Waters’

Skip ‘Ligand Flexibility’

Leave the defaults for ‘Fitness & Search Options’ - it will use CHEMPLP scoring function.

‘GA Settings’ – 10%

Output Options

Output directory: leave as it is (‘.’)

UNCHECK – save ligand rank (.rnk) files

UNCHECK – save ligand log files

UNCHECK – save initialized ligand files

Save solutions to one file:

‘YourTargetvsYourLibraryRun1.sdf’ e.g. “<YOURPROTEINNAME>vsCB1kRun1.sdf”

bestranking\_list\_filename

‘BestYourTargetvsYourLibraryRun1.lst’ e.g. “Best<YOURPROTEINNAME>vsCB1kRun1.lst”

Skip ‘Information in File’

Under the ‘Selecting Solutions’ tab – select

‘Keep the top-ranked solutions for the best ???? ligands only’

This should be 10% of total divided by the number of cores you will use (5 cores)

This ends up being a decimal, then round up to the nearest integer.

(If you didn’t before) - Use the **countsdf.pl** script to figure out the total number of ligands to do this calculation.

Skip GoldMine

Skip Parallel GOLD – we will run in parallel but it will be executed remotely instead of at this console

Skip ‘Constraints’

‘Atom Typing’ - Automatically **set atom and bond types (for the ligand only)**:

Make sure only one box is checked - ‘Ligand’ only

At the top of the page hit Save

Hit ‘Finish’ to save the file

Save GOLD conf file as **gold.conf**

Save protein as <**PDBname>protein.mol2**

Then close GOLD/Hermes

**VERIFY GOLD.CONF**

go back to WinSCP to VERIFY your newly made gold.conf file and MODIFY it in Wordpad

Set Autoscale to **0.1**

cavity\_file = Cavity file name that you made in the Hermes prep - YourLigand.mol2

-may need to add the extension manually if not present

(do this on the actual ligand file and on the line in gold.conf file)

ligand\_data\_file = Reference to ligand file set (the whole path needs to be there).

Number of conformers is ‘10’

Be sure that ‘set\_protein\_atom\_types = 0’

directory = .

concatenated\_output = verify it matches what you entered before

clean\_up\_option save\_best\_ligands = ??? this is 10% of total divided by the number of processors

protein\_datafile = Protein target file name (PDBname\_protein.mol2 – from Hermes prep)

**JOB SCRIPT**

Obtain a **scriptgoldscanshisjob.sh** from the **/home/chem204/scripts** directory and modify it to run for this job.

**TIP**: it helps to have your **gold.conf** file open in a separate window at the same time to look at.

To say:

#!/bin/csh

$GOLD\_DIR/gold\_scan <firstligand> <lastligand> <slice> <PATHTOLIBRARY>.sdf

For each of the things in braces < > You need to enter a value or filename (or the full path).

The last ligand corresponds to how many total ligands are in the file to screen. Your library will have ~5000 + 4 extra ligands

The <slice> is the number of ligands per processor. We will run on 5 processors. So, we will make this slice be the total number of ligands divided by the number of processors.

<PATHTOLIBRARY>:

/home/chem204/2013/<YOURINITIALSFOLDER>/LABVS3/<FILENAME>.sdf

Verify your files with Dr. B or a mentor before running!

BLADE RULES:

You are only allowed to run ONE job at a time.

You are only allowed to use 5 processors unless given permission to use more.

Run the job by executing the script like you did in the last lab.

Verify they are running by using the command to check the status of the queue.

**$qstat**

Use the “Refresh” icon on the SFTP window, to see that GOLD has created some new files in your directory and is running properly. If the PID files are present then it is not done.

**TIP:** if it fails and you need to delete a whole directory – you need to type this to remove it recursively (i.e. with all files inside)

**$rm –r directoryname**

If it does not run, Save the gold.err file and show to a Mentor or Dr. B

**RUN1 ANALYSIS**

After the Run1 completes.

Verify the bestranking.lst has ~1005/5 ligands for each processor

Make note of how long it takes to run by looking at the ‘Date Created’ info for the files

**CONCATENATING OUTPUTS - scriptgoldcat.sh**

Once the job is done, you will need to put the results together for the second run. You can use the **scriptgoldcat.sh** file

You can find this file in the **/home/chem204/scripts** directory

Copy both it over to the directory where you are running your jobs (run 1)

We need to make this have executable permissions. So, type:

$chmod u+x scriptgoldcat.sh

**TIP:** It helps to open your **scriptgoldscanthisjob.sh** and **gold.conf** files and view them while you are trying to type the command

Then edit the command in the script file (replacing what is there already):

/home/chem204/scripts/gold\_cat\_vds ~~1 505 101 Output.sdf~~

To:

/home/chem204/scripts/gold\_cat\_vds <firstligand> <lastligand> <slice> <OUTPUTFILENAME>.sdf

Where the first, last, and slice are the same as what you had for your **scriptgoldscanthisjob.sh** for the first run

**And where** <OUTPUTFILENAME>.sdf is the concatenated\_output from your gold.conf file (instead of the library name)

Be sure to type everything exactly right.

Convert the file to a UNIX file by typing:

#$dos2unix scriptgoldcat.sh

To execute this script type: $sh scriptgoldcat.sh

If it works – you will have a new concatenated file in your **/VS3**

If it fails – make it executable (see the notes in the script file itself)

**VERIFY POSES**

Transfer this new output SDF file to the Desktop. To check that these output poses actually dock in to the active site of the protein, open the original PDB file in PyMol along with these output ligand poses from your Run1.

**Q:** Do the Run1 ligands dock in the active site?

**Q:** Are there the right number of ligands in the Output for Run1?

**CHECKING FITNESS SCORES OF RUN1**

You also want to find out what the Fitness score was for your Extra ligands (pNPP, known inhibitors, etc.) Go into the bestranking lists of this 1st run and search for these values using **Ctrl+F** and the ID number for each compound. HINT: these should have been run on the last processor in your runs.

**1st Run Results for Added Compounds:**

|  |  |  |
| --- | --- | --- |
| **Compound Name** | **Compound PubChem ID** | **GOLD Fitness Score (Run1)** |
| Positive Control 1 |  |  |
| Positive Control 2 |  |  |
| pNPP |  |  |
| Aspirin |  |  |
| Original ligand |  |  |

**RUN2**

**Run secondary gold job**

Create a new directory within the current directory. Name it **Run2**

Copy the **gold.conf** into the new directory

Make changes to the **gold.conf**

Change the Autoscale to **2**.

Be sure the **cavity\_file**  line tells GOLD to look up one directory to find the cavity file.

Change your input ligand file so that it is the ‘**concatenated\_output’** from your first run

IF your **scriptgoldcat.sh** worked, then your ligand output will be in the VS3 directory – it can stay there.

Just make sure the link points to it in the **VS3** directory

Change the concatenated\_output to include ‘**Run2’** in the filename

Change Bestranking to include ‘**Run2’** in the filename

For the ligand solutions, save 10 ligands per processor. This way you will have ~30 in the end for 3 processors.

clean\_up\_option save\_best\_ligands XX

Be sure the **protein\_file** line tells GOLD to look up one directory to find the protein file.

Also, copy the **scriptgoldscanthisjob.sh** to this **Run2** directory and **modify** it for your second run.

Run the 2nd run with the script file.

Make note of how long it takes to run by looking at the ‘Date Created’ info for the files

**RUN2 ANALYSIS**

**Concatenate Output Poses**

After Run2 is done, concatenate the output files.

Move a copy of the **scriptgoldcat.sh** script into the **RUN2** directory.

Then edit the command in the script file (replacing what is there already):

/home/chem204/scripts/gold\_cat\_vds ~~1 505 101 Output.sdf~~

To:

/home/chem204/scripts/gold\_cat\_vds <firstligand> <lastligand> <slice> <OUTPUTFILENAME>.sdf

Where the first, last, and slice are the same as what you had for your **scriptgoldscanthisjob.sh for Run2**

**And where** <OUTPUTFILENAME>.sdf is the concatenated\_output from your gold.conf file of Run2.

Execute the script as you did before. …..

**Q:** Are there the right number of ligands in the Output for Run2?

**CONCATENATE YOUR BESTRANKING FILES - scriptgoldbesrankincat.sh**

You can use the **scriptgoldbestrankingcat.sh**

You can find this file in the **/home/chem204/scripts** directory

Copy it over to the directory where you ran your jobs (run 2)

**NOTE:** This is a little different from the scripts we have been using. You will not edit this file directly, but rather you will feed it different parameters directly on the command line prompt when you execute it in the terminal window.

Type this to execute it:

$sh scriptgoldbestrankingcat.sh <firstligand> <lastligand> <slice> <BESTRANKINGFILENAME>.lst

Where the first, last, and slice are the same as what you had for your **scriptgoldscanthisjob.sh** of Run2.

**And where** < BESTRANKINGFILENAME >.lst is the **bestranking\_list\_filename** from your gold.conf file

**Make a table of the Top 10 ligands + Positive and Negative Controls for your notebook**

**NOTE:** the Positive and Negative Controls may not have made it to the Run2. If so, you will have to get their Bestranking info and Poses from **Run1. (HINT:** use Ctrl+F to find the ligands once you open that Bestranking list**)**

In Excel, use the Text Import Wizard. Open Excel first, and then Open the files (remember to force Excel to ‘All Files’). and then combine data from the jobs for each processor by copying and pasting into one table for the Analysis.

**Sort by Fitness score**.

Add a column called ‘**RANK’** and put the rank of the compound. For the POS and NEG controls – if they were beyond the Top 10 ranking – just put **‘ >10’**

Fix the header row along the ‘Fitness’ line so that the column labels read correctly.

Also, adjust the column width to fit the data.

The GOLD output should show S(PLP) hydrophobic score, S(hbond) hydrogen bond score, S(cho), S(metal), DE(clash), DE(tors), and ‘intcor’ and the ligand identification (a 6-7 digit number). You DO NOT need the File name, or Time shown in your table. Now that you have the Top 30, number them and then you can hide the rest (or delete them). Then save the Excel sheet as a new document so that you don’t copy over your original **bestranking.lst**

**Give your table a super-awesome caption!**

**Lipinski’s Data for Top 10 + Positive and Negative Controls**

Include a 2-D image of the **Top 10** compounds + the **Controls (total of 15)** and the associated Lipinski’s data. You can find this information about the compounds - LogP, number of hydrogen bond donors and acceptors, and the molecular weights - by searching on the ChemBridge website (this is the commercial supplier) using the compound ID (you can enter all at once):

<http://www.hit2lead.com/search.asp?db=SC&SearchPage=id>.

Copy this info into your Excel table to include it for each of the Top 10 + Pos and Neg Ctrls.

The 2-D pictures can be pasted separately from the table but they need to be labeled by compound name/identifier to easily identify which is which.

For the Positive and Negative Controls ligands, you will need to go to **PubChem** website to get the Lipinski’s data since these are not Chembridge ligands. The 2-D pictures can be pasted separately from the table but they need to be labeled by compound name/identifier to easily identify which is which.

Do the Top 5 compounds follow Lipinski’s rule of five? Put a ‘Yes’ or ‘No’ in a column of your Excel table to show this.

**SHOW GOLD POSES FOR THESE 7 LIGANDS:**

* Show the docking poses in the active site of your phosphatase enzyme in a similar fashion as the VDS2 lab (define active site, hide rest of protein, show hydrophobics as sticks, label 4-5 residues). **Zoom in sufficiently as well.** 
  + Top **2** novel ligands from the Chembridge Library
  + **4** Positive Controls (including the Original ligand)
  + **1** Negative Control - aspirin
* **Give your figures super-awesome captions!**

NOTE: if the controls did not make it through to your final docked ligand list due to low scores, you will need to figure out a way to get these poses!

**Polar Contacts**

Find polar contacts in PyMol and list them in the Excel table for these same **6** ligands

**Q:** Did the 5 added compounds (2 known inhibitors of phosphatases taken from the BindingDB.com database, surrogate substrate pNPP, aspirin (acetylsalicylic acid) and Original ligand from PubChem) make it through to show up in the 2nd Run list? If so, where? If not, where do they show up in the 1st Run List. Is this what you expected from these compounds?

**Q**: Are these scores any good? Are you satisfied with the results or should more screening be done against this target with other libraries?

NOTE: if you got an exceptionally high score for a ligand – let Dr. B know. It may be worth pursuing.

**BONUS** (+0.25 points out of 10):

Now that you pretty much know how to set up a screen, take a look at a Blockbuster drug and see if your library has any good hits against the target.

For example,

**Lipitor, Plavix, Seretide, Nexium, Crestor, Seroquel, Humira, Enbrel, Remicade, Zyprexa**

go to: [http://en.wikipedia.org/wiki/Pharmaceutical\_drug#Blockbuster\_drug](http://en.wikipedia.org/wiki/Pharmaceutical_drug%23Blockbuster_drug)

Or, try the targets of **Aspirin**, **Penicillin**, **Erythromycin**, **Ciprofloxacin**, **Viagra** or something else interesting.

Find the real drug name (not the trade name). Then enter this into the **PDB database**. Can you find PDB structures of these blockbuster drugs and their targets?

Choose one and set up a screen against it using your library and the Blockbuster drug as taken from PubChem. Show a bestranking table of the Top 5+Known Drug and Single Best PyMol Image vs. the known drug PyMol Image.

Can you find something that binds better than the existing therapy?